

論点予測に基づく動的仮説スロット生成を用いた対話フレームワークの検討

橋本慧海¹ 長澤史記¹ 白松俊¹

¹ 名古屋工業大学 e.hashimoto.611@stn.nitech.ac.jp

概要

我々は看護師キャリア面談の事前ヒアリングを支援するスロット生成型対話を開発してきた。しかし、LLMによる動的スロット生成は話題が拡散し、質問意図が伝わりにくい課題があった。そこで、本研究ではユーザ応答から次に深掘りすべき論点(将来判断・感情・問題原因)を予測し、その論点を条件としてアブダクションによる仮説スロット生成と質問生成を制御するフレームワークを提案する。ESCの対話戦略ラベル予測モデルを論点予測の代替として用い、GPT-4oで仮説と質問を生成した。予備実験では同一履歴に対し論点条件を切り替えると生成質問の観点が系統的に変化することを確認し、対話の一貫性向上の可能性を示した。

1 はじめに

医療現場では、看護師の離職や配置のミスマッチが継続的な課題であり、個々の看護師の状況や志向を早期に把握し、適切な支援につなげるために、キャリア面談が行われている。我々はこれまでに、面談前に対話システムが看護師からキャリアに関する情報を収集し、看護管理者向けに要点を整理したレポートを生成することで、面談準備を支援するシステムを開発してきた[1]。この事前収集により、面談当日に初めて不満や離職意向が顕在化して看護管理者が対応に苦慮する事態を減らし、面談をより効率的に進めることが期待できる。

システムがユーザの個別状況に適応するため、我々の既存研究ではLLMによる動的スロット生成を導入し、仮説形成(アブダクション)に基づいて新規スロットを生成する枠組みを検討してきた[2]。しかし、実運用を想定した予備の実験では、キャリア支援に直結しにくい細部へスロットが集中することや、質問意図が共有されないまま話題が遷移して対話の自然さおよびレポート有用性が低下すること

が確認された。すなわち、仮説生成プロセスだけでは、深掘りの大まかな方向性を担保できないことが分かった。

そこで本研究では、スロット生成型対話システムに次に深掘りすべき論点を予測するモジュールを組み込み、仮説の方向性を状況に即して制御するフレームワークを提案する。これにより、より重要度の高い論点への集中と、質問遷移の自然さを両立し、面談準備に資する情報を効率的に収集することを目指す。

2 関連研究

2.1 仮説を用いたインタビューシステム

我々の初期の取り組みでは、初期スロットに加え、LLMが対話履歴から仮説を生成し、スロットを動的に生成することで深掘りを試みた。例えば異動を強く希望しているユーザに対し、「人間関係に問題がある」「業務量が多い」などの仮説を作成し、それぞれを質問観点として展開する。

ユーザシミュレータとの実験では、仮説形成ありモデルは固定スロットのみより必須項目のカバー率や質問の深さで優れ、深掘り機構として有効だった。一方、看護師による予備的なロールプレイ実験を行ったところ、「質問の意図が明確」「唐突な話題転換が少ない」といった質問運びの評価は固定スロットが最も高く、仮説を生成する提案手法では「今なぜそれを聞くのか」が伝わりにくい場面が示唆された。

この結果は、提案手法が観点数を増やす一方で、(i)次にどの仮説を優先して掘るか、(ii)次の質問が対話全体のどこに位置づくか、という探索方針がユーザと共有されず、連続性が弱まる可能性を示す。したがって課題は深掘り能力そのものよりも、深掘りの方向(次に掘る論点)の制御と可説明性にあると考えた。

2.2 戦略予想を用いた対話システム

対話の質向上を目的として、対話の次の行動や発話戦略を予測し、その結果で応答生成を制御する研究が行われている [3, 4, 5, 6, 7]. Liu ら [8] が構築した Emotional Support Conversation (ESC) は、各発話に戦略ラベルが付与されており、戦略予測や戦略条件付き生成の評価に広く用いられる。藤田ら [9] は、ESC の戦略ラベルを予測するモデルを応答生成に接続し、自然さの改善を報告した。

これらの研究は支援行動や対話方針といったマクロな制御を主眼としている。反対に、次に何を論点として深掘りするかというミクロな予測・制御は相対的に少ない [10]. ミクロな論点設定は質問の成否に直結し、誤った質問が対話全体の破綻につながりやすいという設計上の難しさがあるためだと考えられる。

しかし、本研究の動的スロット生成は、ある仮説に基づく質問が不適切でも、他の仮説へ切替えや新規仮説の生成によって対話を継続できる。このため、個々の質問表現の当たり外れよりも、深掘りの方向性（論点）が妥当であることが重要となる。そこで本研究では、ESC を用いた戦略予測モデルを構築し、予測された戦略ラベルに対応する論点を設定することで、深掘りの方向性を制御するフレームワークを提案する。

本研究では、設定した論点を仮説生成（スロット生成）の条件として与えることで、仮説スロットと応答が変化することを確かめる。

3 論点予測質問生成フレームワーク

本研究では、仮説の正確さと方向性を分離し、後者を明示的に制御するフレームワークを導入し、聞き取り能力の改善を目指す。本研究における仮説スロットは、原因や結果を一度で特定する推論ではなく、ユーザの自己開示を促し、必要情報へ到達するための問いの足場として位置づける。このとき仮説は完全に正しい必要はないが、ユーザが話しやすい合理的な候補である必要がある。重要なのは仮説の細部よりも、次にどの論点を深掘りするかという方向づけである。本研究では、探索的な試みとして、3つの論点（将来の判断を問う論点、感情を問う論点、問題の原因を問う論点）を用意した。

具体的なシステムの挙動は (1) ユーザの応答を受け取る、(2) 次ターンの論点を推定する、(3) 選択論

今回のフォーカス（将来）

今回あなたが行うアブダクションでは、ユーザの「今後の働き方」「将来の役割・部署」「職場に残るかどうか」といった将来の選択や見通しに関する推測だけを考えてください。

1. Q を作るときは、「この状態が続いたときに、今後どうなりそうか／どうなりたくないか」というユーザの将来に関する見通しや不安を 1 文にまとめる。（ユーザ視点の状態として書くこと）
2. P では、Q を前提に、ユーザが「どのような働き方・役職・キャリア方向を選びたい可能性があるか」について 1~3 個の将来候補（推測）を作る。
3. 新スロットは将来の選択肢や希望に関するものだけを作る。

図 1 ユーザの将来が論点のプロンプト

今回のフォーカス（感情）

今回あなたが行うアブダクションでは、ユーザの「不安」、「喜び」、「モチベーション」といったユーザの感情に関する推測だけを考えてください。

1. Q を作るときは、「この状態が起きたときにどうい感情だったか、今後どのような感情になりそうか」というユーザの感情状態を 1 文にまとめる。（ユーザが感じている・感じそうなことを事実のように書く）
2. P では、Q を前提に、ユーザが「どのような感情になるか」などについて 1~3 個の感情候補（推測）を作る。
3. 新スロットはユーザの感情に関するものだけを作る。

図 2 ユーザの感情が論点のプロンプト

今回のフォーカス（問題の原因）

今回あなたが行うアブダクションでは、ユーザが抱える問題に対する「仕事の忙しさ」、「給与の低さ」、「職場環境の悪化」、「感情の揺れ」といった原因に関する推測だけを考えてください。

1. Q を作るときは、「いまユーザに起きている問題や状態」を 1 文でまとめる。（ユーザの発話内容に基づく事実・状態を書くこと）
2. P では、Q を前提に、その問題が「なぜ起きているのか、なぜそのような状態になったのか」について 1~3 個の原因候補（推測）を作る。
3. 新スロットは問題原因に関するものだけを作る。

図 3 ユーザが抱える問題の原因が論点のプロンプト

点を条件として仮説スロットを生成する、(4) 生成スロットに基づき質問を生成する。この制御により、仮説が外れても質問が妥当な範囲に留まりやすくなり、質問意図の明確化と対話の自然さの改善を期待できる。具体的なプロンプトを図 1-3 に示す。

ESC 対話戦略ラベル予想モデル 本フレームワークを実装するにあたって、理想的にはユーザの応答に対して、システムが次に話すべき論点を予想するモデルが求められる。しかし、本研究ではリソースの関係上、それには至らなかった。そこで、本研究では ESC 対話戦略ラベル予想モデルを論点予想モデルとして代替する。

ESC において定義されている 8 つの対話戦略ラベルに対して、形式的に以下のように対応付ける。

- 将来の判断：Providing Suggestions
- 感情：Restatement or Paraphrasing, Reflection of Feelings, Affirmation and Reassurance,
- 問題の原因：Question
- その他：Self-disclosure, Information, Others

ESC 対話戦略ラベル予想モデルには、ESC (gpt-4.1-mini によって日本語へ翻訳) を用いて gpt-4.1-nano をファインチューニングしたモデルを用いる。

4 予備実験

提案手法に基づく対話システムの実装と評価の予備的な検討として、ユーザ応答に対して、戦略予想モジュールが出力したラベルを用いてスロットを生成し、そのスロットを用いて質問を行うインタビュー対話システムを実装して評価した。我々の既存研究通り、スロット生成と質問生成の LLM には OpenAI GPT-4o を API 経由で使用し、ファインチューニングモデルも API 経由で使用した。まず、本研究では、3 ターンの対話履歴を用意し、それを共通の対話として、ESC のカウンセラー発話と各論点ごとのシステム応答の違いを調べた。

ESC 対話戦略ラベル予想モデル ESC の内 1500 個の対話データを使用し、対話戦略ラベルモデルの予測精度を求めた。本論文で構築した対話戦略予測部の予測精度は 29.8% であり、これは入力条件が同様である藤田ら [9] のシステムの予測精度 31.2%、Tu ら [6] のシステムの予測精度 28.57% と比較して、遜色はない。

5 結果と考察

5.1 各手法との比較

本来、実際のカウンセラー発話と提案フレームワークが生成する発話は一定程度整合していることが望ましい。しかし本研究では、提案フレームワークが論点条件により仮説スロットおよび質問の方向性を制御できるかを定性的に確認することを目的とし、比較例を示すにとどめる。そのため、同一の対話履歴と直近ユーザ発話を入力として固定し、(i) 実カウンセラーの次発話、(ii) (i) と同じラベルに基づく論点による生成結果、(iii) それ以外の論点を強制した生成結果、を比較する。

=== 共通の対話履歴 ===

ユーザ 0: 大丈夫です。去年ほど寒くないのは、少し感謝すべきことですね。

カウンセラー 0: まあ、小さな幸運に感謝しましょうね :)。ところで、ご自身の気分はいかがですか？何か話したいことはありますか？

ユーザ 1: 大学院のことが心配です。

カウンセラー 1: ああ、それはお気の毒に思います。あまり心配しすぎないようにしてくださいね、それはよくある心配事ですから。詳しいことを話してみませんか？

ユーザ 2: もちろんです。私の授業の一つは非常に難しく、教授は全く対応してくれませんが、TA は親切に時間を割いて宿題を手伝ってくれますが、それでも成績はあまり良くありません。そのせいで、自分がそのクラスにいる資格がないように感じて、とても馬鹿に思っています。

それぞれの論点による質問内容の違いを表 1 に示す。実際のカウンセラー発話 (ラベル: Reflection of feelings) は「自分が間違った場所にいると感じているか」「落第してしまうのではないか」といったユーザの自己評価に関する感情の反映・確認であり、感情フォーカスと近い位置づけにある。

次に、表 1 より、論点条件を切り替えることで、生成される仮説スロットおよび質問の観点が系統的に変化することが確認できる。具体的には、感情フォーカス (予測) では「自信の喪失」「ストレスレベル」「モチベーションの低下」など感情状態に

手法	仮説スロット	各スロットに対応する質問
カウンセラー (Reflection of feelings)	—	つまり、自分が間違っただけのように感じているのですか？ クラスに完全に落第してしまうのではないかと感じていますか？
感情フォーカス (予測)	1) 自信の喪失 2) ストレスレベル 3) モチベーションの低下	1) それは大変ですね。じゃあ、その授業の成績が悪いことで、自信を失っていると感じますか？ 2) それは本当に大変ですね。ユーザさんはその授業のことでストレスを感じていますか？ 3) それは大変ですね。じゃあ、大学院の授業が難しいことで、モチベーションが低下していると感じますか？
将来フォーカス (比較のため強制)	1) 授業のサポートを増やす方法 2) 追加のリソースの利用 3) 授業のドロップの検討	1) それは大変ですね。じゃあ、TA 以外に他の学生や学外のリソースを利用してサポートを増やすことを考えたことはありますか？ 2) それは大変ですね。じゃあ、追加のリソースとして、他の参考書やオンラインコースを利用することを考えたことはありますか？ 3) それは大変ですね。じゃあ、その授業をドロップすることを考えたことはありますか？
原因フォーカス (比較のため強制)	1) 教授のサポート不足 2) 授業の難易度: 3) 基礎知識の不足	1) それは大変ですね。教授に直接相談したことはありますか？ 2) それは大変ですね。じゃあ、その授業の内容が特に難しいと感じる部分はありますか？ 3) それは大変ですね。じゃあ、その授業の基礎知識について、事前に十分に学んでいましたか？

表1 同一履歴に対する各論点条件の比較 (対話戦略予測の出力は Reflection of feelings (感情) であった)。

関するスロットが中心となり、質問も「成績が悪いことで自信を失っていると感じるか」「授業のことでストレスを感じているか」といった、感情の言語化・明確化を促す内容になった。

一方、将来フォーカス (強制) では「授業のサポートを増やす方法」「追加のリソースの利用」「授業のドロップの検討」など、将来の選択肢や見通しに関するスロットが生成され、質問も「TA 以外のリソース利用を考えたことはあるか」「授業をドロップすることを考えたことはあるか」といった、将来判断を促す方向へ切り替わった。

また、原因フォーカス (強制) では「教授のサポート不足」「授業の難易度」「基礎知識の不足」など原因候補に関するスロットが生成され、質問も「教授に直接相談したことはあるか」「難しいと感じる部分はどこか」など、原因探索に寄った内容となった。

このことから、予測ラベルは少なくとも当該例において、次発話の方針 (どの観点を深掘りするかの近似として機能していることが示唆される。以上より、本フレームワークは対話履歴に対して「次に深掘りすべき論点」を明示的に切り替え可能であり、質問の意図を一定程度安定化できる可能性が示唆された。

6 結論

本研究では、看護師のキャリア面談の前段階で行う事前ヒアリングを行うスロット生成型対話におい

て、LLM による動的スロット生成が話題拡散や質問意図の不明瞭さを招きやすいという課題に対し、ユーザ応答から「次に深掘りすべき論点」を予測し、その論点を条件として仮説スロット生成と質問生成を制御するフレームワークを提案した。提案フレームワークは、仮説の「方向性」と「正確さ」を分離し、前者 (将来判断・感情・問題原因) を明示的に切り替えることで、質問意図の共有と対話の一貫性向上を狙うものである。

予備実験として、ESC の対話戦略ラベル予測モデルを論点予測の代替として用い、GPT-4o により仮説スロットおよび質問を生成するインタビュー対話システムを実装した。同一の対話履歴に対して論点条件を切り替える比較により、生成される仮説スロットおよび質問の観点が系統的に変化することを確認し、論点条件による質問意図の安定化と対話一貫性向上の可能性を示した。

今後の課題として、現状の論点予測は単一ラベル出力である一方、実対話では複数論点が併存しうるため、多ラベル出力や分布出力に基づく制御が必要である。さらに、単一ターンでは話題の受け渡しが一致しない場合があることから、応答単位の制御に加えて、対話全体で不足している論点を補うようなマクロな論点制御も求められる。

謝辞

本研究は、JST 次世代研究者挑戦的研究プログラム (JPMJSP2112), AIP チャレンジプログラム (JPMJCR20D1) の支援を受けた。

参考文献

- [1] 橋本慧海, 中野幹生, 櫻井崇貴, 白松俊, 駒崎俊剛, 土屋志保. 看護師のキャリア面談のための LLM を用いたスロット生成型対話システム. 人工知能学会研究会資料 言語・音声理解と対話処理研究会, Vol. 100, pp. 198–203, 2024.
- [2] Ekai Hashimoto, Mikio Nakano, Takayoshi Sakurai, Shun Shiramatsu, Toshitake Komazaki, and Shiho Tsuchiya. A career interview dialogue system using large language model-based dynamic slot generation. In **Proceedings of the 31st International Conference on Computational Linguistics**, pp. 1562–1584, 2025.
- [3] Seiya Kawano, Koichiro Yoshino, and Satoshi Nakamura. Neural conversation model controllable by given dialogue act based on adversarial learning and label-aware objective. In **Proceedings of the 12th International Conference on Natural Language Generation**, pp. 198–207, 2019.
- [4] 濱健太, 基村竜晟, 大塚淳史, 中辻真. 発話意図の同時生成によるユーザ情報抽出を目的とした LLM エージェントの対話戦略学習. 人工知能学会研究会資料 言語・音声理解と対話処理研究会, Vol. 102, pp. 115–120, 2024.
- [5] Koji Tanaka, Junya Takayama, and Yuki Arase. Dialogue-act prediction of future responses based on conversation history. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop**, pp. 197–202, 2019.
- [6] Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. MISC: A mixed strategy-aware model integrating COMET for emotional support conversation. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 308–319, 2022.
- [7] Can Xu, Wei Wu, and Yuehua Wu. Towards explainable and controllable open domain dialogue generation with dialogue acts. **ArXiv**, 2018.
- [8] Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. Towards emotional support dialog systems. In **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)**, pp. 3469–3483, 2021.
- [9] 藤田敦也, 上乃聖, 李晃伸. 大規模言語モデルを用いた Emotional Support Conversation システムの構築とその評価. 言語処理学会 第 30 回年次大会 発表論文集, pp. 1378–1383, 2024.
- [10] Xiao-Fei Wen, Wei Wei, and Xian-Ling Mao. Sequential topic selection model with latent variable for topic-

grounded dialogue. In **Findings of the Association for Computational Linguistics: EMNLP 2022**, pp. 1209–1219, 2022.